Aalto University
School of Business

Stock pricing in geographically segmented domestic markets

A replication study on the stock price implications of local bias

Bachelor's Thesis
Henry Keinonen
Aalto University School of Business
Finance
Spring 2019

Author Henry Keinonen

Title of thesis Stock pricing in geographically segmented domestic markets

Degree Bachelor's Degree

Degree programme Finance

Thesis advisor(s) Michael Ungeheuer

Year of approval 2019 **Number of pages** 27 **Language** English

Abstract

In this thesis, I examine the stock pricing implications of local bias – investors tendency to invest in stocks of firms headquartered close to where they live. Theory suggests that the price of a stock should be decreasing linearly in a ratio of aggregate book value of firms in a region divided by the aggregate household income in the same region. By testing this hypothesis on the nine U.S. Census regions, I find support for the theory, results being both statistically, and economically significant. Additionally, local bias treats firms of different size and visibility differently, with the prices of smaller, less visible firms varying more between different regions. Latest data also implies, that the phenomenon persists through time, while the effect of local bias is somewhat reduced in recent years.

Keywords local bias, asset pricing, segmented markets

Contents

Tables

1. Introduction

Individual investors' tendency to invest in domestic stocks instead of foreign ones is one of the better documented biases (French & Poterba, 1991; Tesar & Werner, 1995), with recent studies supporting the persistence of the phenomenon, regardless of decreased foreign investment costs (Levy & Levy, 2014). An extension of the home bias is the local bias – investor's tendency to invest locally inside their domestic market – which has gained more academic interest in the 2000s. The implications of the theory are that, in addition to globally segmented markets, even domestic markets are segmented from an investor's perspective. This phenomenon has been found to apply to both individual investors and professional investment managers and has been studied from perspectives varying from information asymmetries to stock pricing consequences to portfolio returns (Coval & Moskowitz, 1999; Hong, Kubik, & Stein, 2008; Seasholes & Zhu, 2010). Moreover, it has been even discussed in recent studies, that investors tend to invest locally based on their birthplace (Lindblom, Mavruk, & Sjögren, 2018).

Basic asset pricing theory suggest that an independent market's returns should be defined by the systematic risk in that area. Should the market be small enough, in our case "local", the local market conditions and investor risk tolerance should widely determine the risk premium induced in the prices of stocks in the said market. Taking the theory to the extreme and saying that all investors in a country only invest in their hometown stocks would imply that each town would be their own, independent market and that the stocks traded in those markets would only be affected by the local market conditions and local investors' risk tolerance (Hong, Kubik, & Stein, 2008).

This thesis aims to replicate and confirm the stock price effects of local bias described by Hong, Kubik and Stein (The only game in town: Stock-price consequences of local bias, 2008). Their main argument is, that stock prices should be affected by regional risk and risk tolerance and that the stock price consequences should differ for firms traded on the national and local level. To proxy for this effect, the authors construct a variable, RATIO, equal to the aggregate book value of all firms in a region, divided by the aggregate income of all households in the same region. The tests are conducted both on state and U.S. Census region level and results clearly support the hypotheses of the authors – for example, moving from the region with highest RATIO to the region with lowest RATIO, stock prices increase 7.9%. In addition to replicating the results, I aim to test whether the phenomenon persists in recent data and to conduct further robustness checks on the model.

The starting point for my tests is to replicate the values of RATIO on Census region level. I chose not to replicate the results on state level, since state level data ignores too much of the potential investor base of stocks and ultimately results in similar findings as the more broad Census region level (Hong, Kubik, & Stein, 2008). To test the effect of RATIO, I run a cross-sectional regression of log of a firm's market-to-book against RATIO, along with several controls. With three different approaches, the results are similar to the findings of Hong, Kubik and Stein (2008) – the same move from the region with highest RATIO (Middle Atlantic) to the region with lowest RATIO (Deep South) implies a stock price increase of 9.6% in my data.

Consequently, the theory also suggest that the stock price consequences of local bias should differ for large, nationally well-known firms and smaller, locally traded firms. The conclusion is intuitive – large blue-chip companies like Johnson & Johnson, Apple or Coca Cola are largely independent of the local market conditions, regardless of where they are headquartered at. To test this hypothesis, Hong, Kubik and Stein (2008), suggest two proxies for firm visibility – low visibility firms are those that fall outside the top quartile measured by sales and residual of shareholder number against firm sales. Both approaches result in the same conclusion, with RATIO having a bigger impact on smaller, less visible firms. A move from Middle Atlantic to Deep South equals a stock price increase of 2.8% for visible firms and an increase of 9.1% for less visible firms, measured by firm sales. The shareholder number based approach produces an even larger effect, with a 2.9% increase in stock price for visible firms and 12.2% increase for less visible firms.

Finally, I look at the persistence of the phenomenon through time by adding new years to the data. Hong, Kubik and Stein's (2008) data ends at 2005, so adding the most recent accounting data from 2017 extends the time-series by 12 years. To keep the values and coefficients on RATIO comparable, I compute the RATIO in the same way as before and use the same approaches to measure the statistical and economic significance. The results suggest, that the economic magnitude of RATIO has decreased in recent years. A move from Middle Atlantic to Deep South increases stock price by 8.6%, which is 1 percentage point less than with the original data. As a final note, I look at the differences in my replication results and try to find a justification for these differences.

2. Theory and hypotheses

2.1. Model assumptions

Hong, Kubik and Stein (2008, p. 22) make two basic assumptions of a domestically segmented markets. First, a country has *N* regions, with each region having two kinds of firms – visible (V) and hometown (H) firms. Second, each region has two kinds of investors – generalist and local experts. Generalists can invest in visible firms in all N regions, whereas local experts can only invest in hometown firms in their own region. Combining the two assumptions divides the market to *(N + 1)* segments – *N* local ones for each region and one national market for visible firms. The division between firms and investors makes sense considering the findings of Coval and Moskowitz (1999, p. 2047), who argue that locally traded firms (i.e. hometown firms) tend to be smaller, highly levered firms, producing locally traded goods. Kumar and Sulaeman (2015) argue that the geographical dispersion of value relevant information about smaller firms tend to be low and portfolios investing locally generate higher risk-adjusted returns via information asymmetries.

Like Hong, Kubik and Stein (2008, p. 22), I denote each visible firm *i*, in region *j* with F_{ij}^V and each hometown firm with F_{ij}^H respectively. The book values of visible and hometown firms are denoted by B_{ij}^V and B_{ij}^H, with firm dividends at time 1 defined as follows:

$D_{ij,1}^V = r_{ij}^V B_{ij}^V$ for visible firms

$D_{ij,1}^H = r_{ij}^H B_{ij}^H$ for hometown firms

The dividends of each firm can be thought as the return on book equity and are assumed to be normally distributed with means of R_{ij}^V and R_{ij}^H and variances of 1. Additionally, all investors are assumed to have constant-absolute-risk-aversion (CARA) utility (Hong, Kubik, & Stein, 2008). Also known as exponential utility, the CARA utility function implies that changes in risk aversion are independent of changes in wealth as opposed to constant-relative-risk-aversion (CRRA) utility, which assumes risk aversion to change in relation to changes in wealth (Pratt, 1964). The benefit of assuming CARA utility for all investors is the ease of computation and as Hong, Kubik and Stein argue (2008, p. 23), it still allows assuming a correlation between the level of investors income and risk aversion. Based on the CARA utility, the total risk tolerance in a region *j* is denoted by T_j, which is distributed between generalists and local experts by fractions Θ and *(1 - Θ)*. The final assumption is based on the arbitrage-pricing-theory (APT) (Ross, 1976) with only one factor, allowing the assumption that in all regions, there are enough firms to create a diversified portfolio with only market risk. In other words, the one-factor APT is simply the CAPM, with a risk-free rate of zero. Thus, the dividends are perfectly correlated and all hometown firms in a single region can be treated as a single firm.

2.2. Model and hypotheses construction

The effect I am looking at here is the effect of local bias on stock pricing through investor risk tolerance. To establish the existence of such relation and get an idea of the economic magnitude, I use the same model as Hong, Kubik and Stein (2008) to have the same basis for my replication. Furthermore, using the same model allows me to experiment with additional data holding all else equal. For both visible and hometown firms, I am using Tobin's Q - a firm's total market value divided by its total asset value (i.e. book value) - as the primary dependent variable. Treating all hometown firms as one, the market value of this firm in region *j* is defined as:

$$V_j^H = \sum_j \left(R_{ij}^H B_{ij}^H\right) - \frac{\left(\sum_j B_{ij}^H\right)^2}{(1-\theta)T_j}$$

The equation implies that a firm's value is equal to the dividend at time 1 less the variance of the dividend divided by the aggregate risk tolerance of all investors in the region. Dividing the above equation with the aggregate book value of firms in region *j* gives us the market-to-book ratio of a firm *i* in the region:

$$Equation\ 1: Q_{ij}^H = R_{ij}^H - \frac{\sum_i B_{ij}^H}{(1-\theta)T_j}$$

From this equation, I can directly compute the market-to-book ratio for visible firms across all regions. The intuition behind this equation is, that generalists benefit from the returns of all visible firms from all regions, but also have to bear the total risk of all visible firms across all regions, as the returns are perfectly correlated.

$$Equation\ 2: Q_{ij}^{V} = R_{ij}^{V} - \frac{\sum_j \sum_i B_{ij}^{H}}{\theta \sum_j T}$$

As suggested by Hong, Kubik and Stein (2008, p. 22), these equations have two direct implications, which I use two construct my two hypotheses. First, the numerator in equation 1 implies, that the greater the aggregate book value (i.e. risk) of all firms in a region is, the greater is the discount applied to the returns of all firms in a region. Second, equation 2 implies that for all visible firms, the risk premium should be the same as they share the same market as opposed to hometown firms. The consequence of this assumption is, that the effect of local bias on those firms should be zero, as they bear no discount from local market conditions. These two implications lead to two hypotheses and two empirical tests based on them:

Hypothesis 1. For hometown firms: $M/B = \beta_1^H ROE_{ij}^H + \beta_2^H \frac{\sum_i B_{ij}^H}{(1-\theta)T_j}$. The first coefficient should be positive, and the second coefficient should be negative.

Hypothesis 2. For visible firms: $M/B = \beta_1^V ROE_{ij}^V + \beta_2^V \frac{\sum_j \sum_i B_{ij}^H}{\theta \sum_j T}$. The second coefficient for hometown firms should be lower than for visible firms (i.e. $\beta_2^H < \beta_2^V$)

To have an empirical measure for the second variable Hong, Kubik and Stein (2008, p. 23) suggest the variable RATIO, which equals the total book value of all firms in a region divided by the total income of all households residing in the same region.

3. Data and variables

3.1. Sources

The data I use comes from the same sources as used by Hong, Kubik and Stein to enable as close replication of results as possible. In addition to replicating the original dataset, I also extend the time period of the data from the original 1970 to 2005 to 1970 to 2017. To compute the RATIO variable, I obtain data on personal income from the database of the Bureau of Economic Analysis (BEA). The BEA database is also used for obtaining data on other demographic variables, such as population density. Firm level market and accounting data is obtained from the Center for Research in Security Prices (CRSP) and Compustat respectively. Market data from CRSP includes stock prices and shares outstanding for companies as well as the exchange and industry classification codes. Accounting data from Compustat includes book value of equity, depreciation, net income, sales, research and development expenditure and state of headquarters.

3.2. Main variables

Using year-end data on market equity value and book equity value, I compute the market to book -value for each firm in the dataset. Following the procedures of Hong, Kubik and Stein (2008, pp. 25-26), I take the natural logarithm of this variable to make the variable more robust in terms of variation, resulting in a more symmetrical distribution. This variable, ln(M/B), is my main dependent variable. The main explanatory variables in my regression are the RATIO, return on equity and R&D to sales. Also included in the regression are a dummy for whether the firm reports R&D expenditure, a dummy for exchange listing (1, 2 and 3 for NYSE, AMEX and Nasdaq) and a dummy of four-digit SIC-codes for industry classification.

The RATIO variable is computed from BEA data on personal income and Compustat data on equity book value as the ratio of aggregate book value in a region in year t to aggregate income in the same region and year. In calculating the RATIO, I exclude dividend income from the aggregate income in a region to avoid a straight relationship between stock prices and RATIO, as suggested by Hong, Kubik and Stein (2008). Additionally, the RATIO variable is recalculated for each firm, excluding the specific firm's market book value from the computation, as otherwise the same variable would be on both sides of the regression. Return on equity (ROE) is defined as a firm's net income divided by previous year book equity value. R&D to sales is a firm's research and development expenditure divided by same year sales. To factor in outliers in the main explanatory variables and make them more robust, the ln(M/B), ROE and R&D to sales variables are winsorized at the 1% and 99% levels.

3.3. Control variables

In addition to the baseline specification, I add several controls to the regression in further robustness checks. State level population and population density are both obtained from the BEA database, which I use to compute Census region level per capita income and population density. I use three future variables for each company – future region income growth, future firm ROE and future firm sales growth. The income and sales growths are defined as the average growth rate between years $t + 1$ and $t + 3$, whereas future ROE is a simple average of the firm's ROE between years $t + 1$ and $t + 3$.

Industry codes are used to identify conglomerates and firms operating in dominant industries in their region. Firms segment information is available from 1976 onwards and I use this information to compute a conglomerate dummy for firms operating in multiple segments. Dominant industries are identified as those, which contribute to more than 10% of the aggregate book equity value in a region in a year and those industries are dropped from the regressions. Finally, I also add the natural logarithm of sales and an S&P 500 dummy for firms in the S&P 500 index.

3.4. Replication of RATIO

To replicate the results of Hong, Kubik and Stein (2008), it is important to match the raw data as close as possible before running the regressions. The only raw data available as reference is the RATIO, but luckily this is also the variable I am most interested in. Table 1 presents the RATIO on Census region and state level for every five years from 1970 to 2005. In addition, time-series and cross-sectional averages and standard deviations are presented. The results are consistent with the reference values of Hong, Kubik and Stein (2008, pp. 27-28). Middle Atlantic is the highest ranking of all Census regions, with RATIO averaging 0.66 (0.77)[1] and Deep South ranking the lowest with an average RATIO of 0.16 (0.21)[1]. The

[1] Table 3 (Hong, Kubik, & Stein, 2008, p. 27)

rankings of the Census region RATIOs are also consistent through time. Regressing the rankings of the RATIOs on their lagged values results in a coefficient of 0.99.

We can also observe the same anomalies in the state level RATIOs as in the results of Hong, Kubik and Stein (2008, pp. 27-28). Throughout the time-series, Nebraska has had the highest average RATIO with largest standard deviation, which is largely due to the presence of Berkshire-Hathaway. Similar high RATIO anomalies also apply to Arkansas, with the presence of Wal-Mart and Delaware due to tax laws favourable to businesses. These kinds of outliers would imply a measurement error in the RATIO, as the variable only catches the effect of a firm's headquarters and not the actual area of operations (Hong, Kubik, & Stein, 2008).

The replicate average RATIOs, however, seem to be lower than the reference values, both on Census region and state level. To address this issue and further confirm a successful replication, I run a regression between Hong, Kubik and Stein's RATIOs and my RATIOs. On state level, the regression is run on time-series and on Census region level, both on time-series and cross-sectional regression as Census regions are the focus regions. The outputs of the regression are displayed on Table 2, presenting the regression coefficients, R-squareds and the time-series and cross-sectional averages along with standard deviations. We can see that the time-series regressions produce almost identical results, regardless of the level of region, with both coefficients averaging at 1.12. The cross-sectional regression also produces good results, with an average coefficient of 1.00. Even with this limited dataset, I can safely assume that the replication also works for all years from 1970 to 2005 and that the regression replication should also work.

My earlier concern of lower RATIOs is further confirmed by the regression results, with time-series coefficients averaging over 1.00. As the RATIO values are consistently above 1.00 however, this should not be an issue. Another remark is to be made from how the RATIO acts on a cross-sectional regression. Even though the average coefficient is exactly 1.00 and the coefficients are dispersed more evenly around 1.00, there is more deviation in the data and the R-squareds for Midwest and Mountain regions drop significantly lower than the R-squareds of other regions. This is an expected result as there is likely to be more variation cross-sectionally than through time. One explanation for such behaviour is that Compustat only has the most recent headquarter state of each firm. What this means is that RATIO doesn't capture the information when a company moves its headquarters to another state. This could also explain the difference in the cross-sectional data, as the reference data is from 2005 and my replication is based on latest data from December of 2017.

3.4. RATIO with additional years and new hypothesis.

Next, my attention turns to computing the RATIO on Census region and state level with completely new data. With the original data ranging from 1970 to 2005 and latest data ranging up to 2017, I have 12 additional years to do the tests with. In computing the new values of RATIO, I focus entirely on Census regions and comparing the new values to old ones. The results of this comparison are displayed on Table 3, presenting the time-series means and standard deviations of RATIO on three different timelines – original from 1970 to 2005, extended from 1970 to 2017 and new from 2006 to 2017. The message of this table is clear; the values of RATIO have increased throughout Census regions. The average increase in RATIO comparing the original and extended data is equal to 10.7%, whereas comparison of the original and new data results in an average increase of 42,9% between Census regions.

Considering the decomposition of RATIO (aggregate book-value / aggregate income in a region), the only sensible explanation is, that the aggregate book-value has increased more than the aggregate income, as it is highly unlikely that either of the variables would have decreased throughout the year. Indeed, this is the case – on average, aggregate book-value has grown by 7,8% in all Census regions, whereas aggregate income has increased on average 6,5%.

Arguably, the increase in RATIO in recent years is radical and raises the question of its economic significance. Looking back at the local bias model, an increase in RATIO keeping all else equal, should result in a lower stock price. Interestingly, the other explanatory variable, firm ROE, has decreased on average by 1% through the whole time period, and while this estimate is purely directional, it gives some reference for the change in RATIO. With stock prices and RATIO increasing at the same time, an explanation emerges endogenously from the model:

Hypothesis 3: For both hometown firms and visible firms, the effect of RATIO has decreased through 2006 to 2017.

Table 1 - Summary of RATIOs on Census region and state level

	1970	1975	1980	1985	1990	1995	2000	2005	Mean	S.D.
Panel A: Census regions										
New England	0.38	0.49	0.52	0.45	0.42	0.51	0.56	0.62	0.49	0.06
Middle Atlantic	0.64	0.69	0.70	0.54	0.48	0.55	0.80	0.93	0.66	0.13
Midwest	0.49	0.51	0.50	0.48	0.45	0.46	0.44	0.48	0.47	0.03
Plains	0.22	0.26	0.28	0.27	0.26	0.35	0.44	0.60	0.32	0.09
Atlantic Coast	0.32	0.34	0.31	0.30	0.27	0.32	0.40	0.47	0.34	0.05
Deep South	0.08	0.13	0.13	0.14	0.13	0.22	0.23	0.24	0.16	0.05
Southern Plains	0.60	0.62	0.60	0.52	0.46	0.49	0.74	0.74	0.58	0.08
Mountain	0.23	0.25	0.20	0.17	0.15	0.24	0.28	0.26	0.22	0.05
West Coast	0.24	0.27	0.27	0.25	0.23	0.30	0.50	0.50	0.31	0.09
Cross-sectional mean	0.36	0.40	0.39	0.35	0.32	0.38	0.49	0.54		
Cross-sectional S.D.	0.19	0.19	0.19	0.15	0.14	0.12	0.19	0.22		
Panel B: States										
New England										
Connecticut	0.69	0.89	0.94	0.67	0.60	0.64	0.48	0.56	0.70	0.16
Maine	0.04	0.04	0.04	0.04	0.05	0.07	0.03	0.02	0.05	0.01
Massachusetts	0.33	0.43	0.45	0.48	0.46	0.61	0.82	0.92	0.56	0.16
New Hampshire	0.16	0.19	0.21	0.13	0.08	0.18	0.10	0.13	0.14	0.04
Rhode Island	0.17	0.23	0.24	0.27	0.30	0.33	0.39	0.45	0.30	0.07
Vermont	0.01	0.03	0.03	0.03	0.04	0.08	0.07	0.06	0.05	0.03
Middle Atlantic										
New Jersey	1.06	1.00	0.97	0.50	0.39	0.56	0.97	0.57	0.74	0.25
New York	0.58	0.67	0.71	0.69	0.64	0.69	1.00	1.43	0.78	0.24
Pennsylvania	0.46	0.51	0.47	0.32	0.27	0.30	0.29	0.38	0.37	0.10
Midwest										
Illinois	0.70	0.74	0.80	0.79	0.71	0.74	0.65	0.69	0.73	0.04
Indiana	0.11	0.14	0.16	0.11	0.11	0.14	0.18	0.38	0.16	0.06
Michigan	0.69	0.63	0.51	0.52	0.53	0.44	0.39	0.29	0.49	0.13
Ohio	0.38	0.43	0.43	0.41	0.33	0.41	0.45	0.51	0.42	0.05
Wisconsin	0.16	0.23	0.21	0.19	0.21	0.24	0.21	0.28	0.22	0.03
Plains										
Iowa	0.11	0.12	0.14	0.15	0.16	0.21	0.08	0.20	0.16	0.04
Kansas	0.08	0.12	0.12	0.12	0.11	0.13	0.28	0.71	0.17	0.11
Minnesota	0.34	0.42	0.45	0.40	0.32	0.41	0.47	0.67	0.43	0.07
Missouri	0.27	0.34	0.36	0.34	0.32	0.40	0.38	0.36	0.35	0.03
Nebraska	0.26	0.25	0.25	0.31	0.38	0.74	1.55	1.96	0.67	0.58
North Dakota	0.14	0.12	0.15	0.08	0.09	0.04	0.09	0.11	0.10	0.03
South Dakota	0.02	0.04	0.08	0.03	0.07	0.12	0.14	0.07	0.06	0.04

	1970	1975	1980	1985	1990	1995	2000	2005	Mean	S.D.
Atlantic Coast										
Delaware	1.59	1.89	1.41	1.70	1.58	0.97	1.07	0.50	1.44	0.39
District of Columbia	0.27	0.36	0.44	0.47	0.64	1.37	1.16	1.60	0.76	0.43
Florida	0.20	0.21	0.17	0.13	0.10	0.15	0.20	0.21	0.17	0.04
Georgia	0.35	0.40	0.37	0.49	0.41	0.44	0.60	0.66	0.46	0.10
Maryland	0.20	0.21	0.21	0.19	0.17	0.20	0.19	0.25	0.20	0.02
North Carolina	0.34	0.38	0.33	0.29	0.26	0.40	0.66	0.93	0.43	0.19
South Carolina	0.14	0.16	0.12	0.09	0.08	0.11	0.10	0.08	0.11	0.03
Virginia	0.60	0.59	0.60	0.52	0.54	0.53	0.56	0.67	0.56	0.05
West Virginia	0.00	0.01	0.00	0.00	0.04	0.04	0.03	0.06	0.02	0.02
Deep South										
Alabama	0.04	0.06	0.07	0.10	0.11	0.16	0.23	0.23	0.12	0.07
Kentucky	0.11	0.16	0.14	0.12	0.11	0.17	0.10	0.14	0.13	0.02
Mississippi	0.00	0.03	0.02	0.03	0.04	0.07	0.06	0.05	0.04	0.02
Tennessee	0.15	0.23	0.22	0.22	0.19	0.35	0.38	0.39	0.27	0.08
Southern Plains										
Arkansas	0.05	0.07	0.10	0.17	0.32	0.54	0.77	1.07	0.36	0.31
Louisiana	0.10	0.11	0.12	0.30	0.31	0.20	0.57	0.16	0.21	0.11
Oklahoma	0.12	0.18	0.20	0.15	0.11	0.17	0.26	0.39	0.19	0.07
Texas	0.91	0.92	0.87	0.68	0.57	0.60	0.83	0.87	0.77	0.13
Mountain										
Arizona	0.33	0.33	0.24	0.21	0.16	0.18	0.19	0.20	0.23	0.06
Colorado	0.23	0.32	0.25	0.19	0.18	0.39	0.51	0.45	0.32	0.14
Idaho	0.22	0.24	0.23	0.21	0.20	0.22	0.42	0.37	0.26	0.07
Montana	0.05	0.06	0.05	0.07	0.07	0.08	0.08	0.02	0.06	0.02
Nevada	0.19	0.20	0.14	0.17	0.15	0.27	0.24	0.29	0.20	0.05
New Mexico	0.00	0.02	0.05	0.07	0.05	0.07	0.08	0.07	0.06	0.03
Utah	0.51	0.42	0.31	0.21	0.17	0.27	0.15	0.19	0.26	0.11
Wyoming	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01
West Coast										
California	0.27	0.30	0.30	0.27	0.24	0.32	0.50	0.51	0.33	0.08
Oregon	0.07	0.15	0.17	0.17	0.21	0.25	0.17	0.17	0.17	0.04
Washington	0.11	0.16	0.15	0.14	0.17	0.20	0.64	0.58	0.25	0.18
Cross-sectional mean	0.28	0.32	0.31	0.28	0.27	0.33	0.40	0.45		
Cross-sectional S.D.	0.31	0.34	0.29	0.29	0.27	0.27	0.34	0.42		

Table 2 - Results of regression reference vs. replicate RATIOs

STATE			CENSUS REGION					
Time-series			Time-series			Cross-section		
	RATIO	R-squared		RATIO	R-squared		RATIO	R-squared
1970	1.13	0.91	1970	1.19	0.96	Atlantic Coast	1.03	0.86
1975	1.15	0.95	1975	1.20	0.98	Deep South	1.06	0.96
1980	1.21	0.94	1980	1.21	0.98	Middle Atlantic	0.95	0.98
1985	1.12	0.93	1985	1.09	0.94	Midwest	0.84	0.57
1990	1.11	0.91	1990	1.09	0.96	Mountain	1.01	0.42
1995	1.18	0.93	1995	1.19	0.96	New England	0.93	0.96
2000	1.18	0.89	2000	0.94	0.92	Plains	1.03	0.96
2005	0.92	0.72	2005	1.07	0.97	Southern Plains	0.91	0.89
						West Coast	1.21	0.99
Mean	1.12		Mean	1.12		Mean	1.00	
S.D	0.09		S.D.	0.09		S.D.	0.10	

Table 3 - RATIO summary on different timelines (Census)

	1970-2005		1970-2017		2006-2017	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
New England	0.49	0.06	0.54	0.09	0.66	0.04
Middle Atlantic	0.66	0.13	0.73	0.17	0.94	0.06
Midwest	0.47	0.03	0.48	0.03	0.49	0.04
Plains	0.32	0.09	0.40	0.16	0.64	0.08
Atlantic Coast	0.34	0.05	0.36	0.06	0.42	0.04
Deep South	0.16	0.05	0.18	0.05	0.21	0.01
Southern Plains	0.58	0.08	0.63	0.12	0.80	0.05
Mountain	0.22	0.05	0.23	0.05	0.27	0.03
West Coast	0.31	0.09	0.38	0.16	0.60	0.08

4. Results

4.1. Regression model and methodology

To test hypothesis 1, the following equation is computed to use as the baseline regression:

$$\ln(M/B) = B_1 ROE + B_2(R\&D/Sales) + B_3 RATIO + B_4 R\&D\_reporting + B_5 SIC + B_6 Exchange$$

The regression follows directly from the earlier model, where market-to-book value is explained by firm ROE and region RATIO, now also including the ratio of research and development expenditure to sales. In addition, the baseline regression includes dummies for whether the firm reports R&D expenditures (1 or 0), a 4-digit SIC code specifying the firm's industry and a dummy for exchange listing, with possible values of 1, 2 and 3 for NYSE, AMEX and Nasdaq. Excluded from the regression are companies with a book value below ten million dollars and companies with a SIC code starting with 6 belonging to financial-services industry. Additionally, RATIO is calculated separately for each firm, excluding the book value of the firm in question to avoid the same value appearing on both sides of the regression.

While ROE and R&D / Sales are calculated separately for each firm, RATIO only has nine independent observations per year, when looking at RATIOs at the Census region level, as noted by Hong, Kubik and Stein (2008, p. 29). To put this into scale, each year has on average a total of 3394 observations. Due to the low number of observations, RATIO is expected to display high cross-correlation. As noted before, the ranks of RATIO vary little over time, so the sample should also display autocorrelation to some extent. To address the issues of cross-correlation and autocorrelation, Hong, Kubik and Stein (2008) propose three different approaches.

The first approach is the Fama-Macbeth regression (Fama & MacBeth, 1973). In this approach, I run a separate cross-sectional regression for each year from 1970 to 2005 making the total number of regressions 36. The coefficients of these regressions are then averaged, and the statistical significance is evaluated based on the standard errors. I use Newey-West adjusted standard errors, with one lag, to correct the coefficients for autocorrelation (Newey & West, 1987).

Fama-Macbeth approach works well when there is no fixed or slow-decaying effect in the data, but understates the standard errors in the presence of the aforementioned, even with Newey-West adjusted standard errors (Petersen, 2009). The second approach, a single pooled regression with clustered standard errors, however, works well even with fixed or slow decaying effects. In this approach, the whole dataset is pooled and all other variables, other than RATIO, are allowed to interact with a year dummy (i.e. the effect of RATIO is fixed). Whereas the Fama-Macbeth approach used Newey-West adjusted standard errors, in single-pooled regression the standard errors are clustered at the region level to adjust for autocorrelation.

As the third approach, Hong, Kubik and Stein (2008) derive a "collapsed" regression from the baseline regression. In this approach, I run the baseline regression for each year, but drop the RATIO from the right-hand side. From these regressions, the residuals are averaged at the Census region level for each year, resulting in nine independent residuals per year. Finally, I run a pooled regression, where the dependent variable is the firm-level residual and the explanatory variables are the RATIO and a year dummy. Similarly, to the single pooled regression, the standard errors are clustered at the region level.

4.2. Baseline replication

The results of the baseline specification with three different approaches, replicating the results of Hong, Kubik and Stein (2008, p. 29), are displayed on Table 4. First, are presented the yearly cross-sectional coefficients on RATIO, R&D / Sales and ROE and R-squared of the Fama-Macbeth regression. Second, are the average coefficients, their statistical significances (*** = 0.001 and ** = 0.01), Newey-West adjusted standard errors and the number of cross-sectional coefficients with the sign predicted by Hypothesis 1. Finally, the coefficients on RATIO are presented for the single pooled and collapsed regressions with statistical significances and clustered standard errors.

Based on the regression coefficients and standard errors, the replication of the effect of RATIO is successful with an average coefficient of -0.153 (-0.150)[2] and a standard error of 0.025 (0.020)[2]. The number of RATIOs with negative sign is 33/36 (35/36)[2], which is still in line with Hypothesis 1, with positive RATIO coefficients only breaking slightly above zero. On the other hand, the replication of the effect of R&D / Sales and ROE is not as successful.

[2] Table 5 (Hong, Kubik, & Stein, 2008, p. 31)

The average coefficients of 3.11 (2.08)[3] and 0.624 (1.79)[3] and the standard errors of 1.02 (2.73)[3] and 0.194 (1.19)[3] deviate too much from the reference values of Hong, Kubik and Stein (2008), to consider the replication of these variables successful. This applies especially to the standard errors, being two times lower for R&D / Sales and six times lower for ROE even with the Newey-West adjustment. The issue is most probably generated by the data itself as the replicate values of RATIO are very close to the reference values. As the effect of the two variables still supports Hypothesis 1, the results, however, do not contradict the original findings of Hong, Kubik and Stein (2008).

With my focus on the RATIO variable, we can now look at the replication of the two remaining approaches – the single pooled and collapsed regressions. Both the pooled regression and the collapsed regression produce results similar to the Fama-Macbeth regression, taking a negative sign as predicted. The results also support the argument of Petersen (2009), with both standard errors increasing compared to the Fama-Macbeth regression. The pooled regression results in a coefficient of -0.183 (-0.136)[3] and a standard error of 0.041 (0.034)[3], while the collapsed regression results in a coefficient of -0.098 (-0.105)[3] and a standard error of 0.028 (0.029)[3]. Pooled regression taking the largest negative coefficient is an unexpected result as it overstates the coefficient on RATIO compared to the other two approaches.

Considering the results of the baseline replication, the economic significance of RATIO should also be closely related to the economic significance that Hong, Kubik and Stein (2008) find in their results. Using the pooled regression as a benchmark, we can look at the difference in RATIO between the highest and lowest values. These RATIOs are Middle Atlantic's 0.66 and Deep South's 0.16. The regression results suggest that the difference in ln(M/B) between the two Census regions is $0.183 * (0.66 - 0.16) = 0.092$. In absolute magnitude, the difference is equal to a 9.6% increase, when moving from Middle Atlantic to Deep South $(e^{0.092} - 1)$. The same effect, that Hong, Kubik and Stein (2008) observe in their results, is equal to 7.9%, leading the replicated pooled regression to overstate the economic magnitude of RATIO.

[3] Table 5 (Hong, Kubik, & Stein, 2008, p. 31)

Table 4 - Summary of baseline specification results on Census region level

	RATIO	R&D to sales	ROE	R-squared
1970	0.037	8.88	3.62	0.41
1971	-0.054	10.7	2.62	0.30
1972	-0.099	12.4	0.212	0.10
1973	-0.087	12.6	3.31	0.27
1974	-0.074	8.27	0.484	0.09
1975	-0.117	9.58	2.03	0.29
1976	-0.127	8.61	2.48	0.34
1977	-0.075	6.76	1.90	0.28
1978	-0.179	7.69	0.805	0.17
1979	-0.163	10.6	0.246	0.11
1980	-0.082	8.73	1.30	0.23
1981	-0.118	4.42	0.181	0.07
1982	-0.266	1.21	0.240	0.09
1983	-0.116	0.173	0.816	0.14
1984	-0.195	0.268	0.481	0.08
1985	-0.193	0.452	0.016	0.04
1986	-0.209	0.283	0.035	0.02
1987	-0.283	0.076	0.082	0.03
1988	-0.215	0.033	0.526	0.07
1989	-0.299	0.034	0.092	0.03
1990	-0.373	0.111	0.073	0.04
1991	-0.340	0.001	0.182	0.06
1992	-0.124	0.003	0.049	0.05
1993	-0.192	0.003	-0.005	0.05
1994	-0.349	0.000	0.005	0.09
1995	-0.407	0.002	0.133	0.12
1996	-0.362	0.001	0.003	0.10
1997	-0.062	0.003	0.051	0.06
1998	-0.153	0.002	0.047	0.08
1999	-0.036	0.004	0.002	0.15
2000	0.032	0.000	0.000	0.08
2001	0.031	0.000	0.253	0.10
2002	-0.047	0.000	0.008	0.04
2003	-0.044	0.004	0.027	0.12
2004	-0.047	0.001	0.050	0.07
2005	-0.114	0.000	0.098	0.08
Avg. coefficient	-0.153***	3.11**	0.624***	
F-M std. error	(0.025)	(1.02)	(0.194)	
# With predicted sign	33/36	36/36	35/36	
Pooled regression	-0.183*** (0.041)			
Collapsed regression	-0.097*** (0.028)			

4.3. Control replication

In additional robustness checks, I replicate the controls of Hong, Kubik and Stein (2008). The summary of these controls is displayed on Table 5, presenting the coefficients on RATIO and each control with their standard errors and statistical significances (*** = 0.001 and ** = 0.01). The baseline specification, presented on row 1, is the earlier pooled regression model, being more robust to the evident autocorrelation in the data. Row 2 adds per capita income in a Census region to the regression. With both the coefficient and standard error approaching zero, the variable fails in adding anything to the regression. One possible explanation for the issue is, that per capita income is measured in absolute value, while the dependent variable is in logarithmic scale. Indeed, per capita income in 2005 is, on average, 8.5 times higher than in 1970. Takings logs of the per capita income values results in a coefficient of 0.102 for per capita income and -0.209 for RATIO and standard errors of 0.034 and 0.042 respectively, with both variables being statistically significant at the 1% level.

Row 3 adds region population density to the regression. The economic significance of this variable is arguably very low, with a coefficient of -0.0003. Similar to the results of Hong, Kubik and Stein (2008), population density effectively invalidates RATIO with a coefficient of -0.007 (-0.003)[4]. The authors address this issue by suggesting that population density better captures the locality of firms, for example through operational presence. The effect of the variable, however, motivates the next three controls as population density might proxy for other variables that actually belong in the regression (Hong, Kubik, & Stein, 2008).

Next, I add future region income growth, future firm ROE and future firm sales growth, both separately and together, on row 3 to 6. Income growth is defined as the average growth rate of Census region aggregate income through years *t + 1* to *t + 3*. Firms sales growth is defined in a similar way for firm sales. Future firm ROE is simply an average of the firms ROE through years *t + 1* and *t + 3*. Adding these variables yields some interesting results – while future ROE and sales growth have little impact on RATIO, future income growth on the other hand affects RATIO more than any other control in the tests. The results also directly contradict the findings of Hong, Kubik and Stein (2008), with income growth taking a coefficient of 2.47 (0.031)[4], implying that income growth in a region actually does have a role in explaining stock pricing.

[4] Table 6 (Hong, Kubik, & Stein, 2008, p. 32)

The rest of controls through rows 8 to 11 are on line with the results of Hong, Kubik and Stein (2008)[4]. Dominant industries on row 8 are defined as those, that constitute to more than 10% of the aggregate book value on state level, in a year. The reasoning behind this variable is, that dominant industries might introduce idiosyncratic risk premium in a local market, otherwise defined by the market risk (Hong, Kubik, & Stein, 2008). Conglomerates on row 9 are defined as firms operating on multiple segments and the two last controls are the log of sales and a dummy for firms in the S&P 500 index. None of these controls affect the coefficient on RATIO in any significant way.

Even with a higher coefficient on RATIO, the effect of adding controls is similar to the results of Hong, Kubik and Stein (2008), leading me to conclude that the difference in the coefficient on RATIO is merely a question of different level. Adding future income growth, however, yields contradicting results both on the coefficient on the variable itself and its effect on the coefficient on RATIO. This is a matter I will look more closely into, later in my additional controls.

Table 5 - Summary of pooled regression with controls

	RATIO	Future region income growth	Future firm ROE	Future firm sales growth	Misc.
1.Baseline specification	-0.183***				
	(0.041)				
2. Add region per capita income	-0.173***				1.9e-6
	(0.043)				(2.3e-6)
3. Add region population density	-0.007				-0.0003***
	(0.055)				(7.3e-5)
4. Add future region income growth	-0.123**	2.47***			
	(0.042)	(0.213)			
5. Add future firm ROE	-0.195***		0.504***		
	(0.040)		(0.090)		
6. Add future firm sales growth	-0.183***			6.5e-5	
	(0.041)			(5.9e-5)	
7. Add all future controls	-0.144***	2.10***	0.498***	8.4e-5	
	(0.040)	(0.217)	(0.089)	(5.7e-5)	
8. Remove dominant industries	-0.174***				
	(0.044)				
9. Add conglomerate dummy	-0.182***				0.019
	(0.041)				(0.017)
10. Add log sales	-0.182***				0.037***
	(0.041)				(0.004)
11. Add S&P 500 indicator	-0.183***				0.335***
	(0.040)				(0.017)

4.4. RATIO and visible vs. hometown firms

Hypothesis 2 assumes that the coefficient on RATIO is dependent on the visibility of the firm – specifically it should be lower for hometown firms than for visible firms. To test this hypothesis, I use the same proxies for firm visibility as Hong, Kubik and Stein (2008). The first proxy is a dummy equalling 1 for firms with sales below the top quartile in any period. The second one is a dummy equalling 1 for firms, with shareholder base falling outside the top quartile in any period. In detail, the second dummy is defined through a regression of the log of a firm's shareholder number against the log of its sales. The residuals of firm shareholders are then used to find the firms falling outside the top quartile.

Next, I run the baseline pooled regression, now also including the dummy for firm visibility and an interaction term between the visibility dummy and RATIO (i.e. $Visibility\ dummy * RATIO$), with both proxies for visibility. Table 6 displays the results of this regression, presenting the coefficients, standard errors and statistical significances (** = 0.01) of RATIO and the interaction term. We can clearly see, that there is no difference on the coefficient on RATIO between the two proxies, with both taking negative signs. The first proxy is closer to the results Hong, Kubik and Stein (2008), with coefficients of -0.056 (-0.071)[5] on RATIO and -0.118 (-0.097)[5] on the interaction term. The corresponding coefficients on the second proxy are -0.057 (-0.056)[5] and -0.173 (-0.109)[5].

The economic significance of visibility is clear, if we look at the difference in RATIOs between Middle Atlantic (0.66) and Deep South (0.16). Starting with the sales proxy, for visible firms, a change from Middle Atlantic to Deep South is equal to a 2.8% increase in stock price. For low visibility firms, the same effect is equal to an increase of 9.1%[6] in stock price. Using the shareholder proxy, the economic significance is even larger. Moving from Middle Atlantic to Deep South is equal to a 2.9% increase in stock price for visible firms and 12.2% for low visibility firms.

Table 6 - The effect of firm visibility on RATIO

Low visibility proxy	RATIO	(Low visibility) x RATIO
1. Sales below top quartile	-0.056 (0.052)	-0.118** (0.052)
2. Residual no. of shareholders below top quartile	-0.057 (0.049)	-0.173** (0.033)

[5] Table 9 (Hong, Kubik, & Stein, 2008, p. 35)
[6] $Exp\big((0.056 + 0.118) * 0.50\big) - 1$

4.5. Tests with additional years

Having confirmed stock pricing implication of RATIO and how it affects visible and hometown firms differently, I now look at the persistence of the phenomenon through time. Hypothesis 3 suggests, that as stock prices and values of RATIO have simultaneously increased, the effect of RATIO has decreased on the contrary. To test my hypothesis, I run the same three baseline regressions on two different timelines – from 1970 to 2017 and from 2006 to 2017. It is noteworthy, that since Fama-Macbeth regression essentially treats each year as one observation and the pooled and collapsed regressions fix the effect of RATIOS, it is highly unlikely that any of the approaches produce any meaningful results on the latter timeline with only 12 independent observations.

Table 7 displays the summary statistics of the three regressions on two timelines, presenting the coefficients, standard errors and statistical significances (*** = 0.001, ** = 0.01) of each variable. The results from the first timeline are very similar to the original regression with coefficients on RATIO taking negative signs as predicted and the ranks of the coefficients still in the same order. The coefficients on RATIO from all three regressions also support Hypothesis 3, with all three coefficients increasing. In absolute values, both the Fama-Macbeth coefficient and the pooled coefficient decrease by approximately 18%, whereas the collapsed coefficient drops by 40%. Considering the economic magnitude of this change, a firm moving from Middle Atlantic to Deep South leads to an 8.6%[7] increase in stock price, whereas the increase was equal to 9.6% in the original sample.

As predicted, the results from the latter timeline are not of much interest. While two of the three regressions retain the negative coefficient on RATIO, all three regressions produce results, that are not statistically significant. It is clear that the newly introduced data affects how RATIO behaves on the whole timeline, but based on these results, it is still too early to say, whether the change in the economic magnitude is permanent or not.

[7] $Exp((0.73 - 0.18) * 0.150) - 1$

Table 7 - Summary of baseline specification with additional years

1970 - 2017	RATIO	R&D to sales	ROE
Avg. coefficient	-0.125***	2.33**	0.483***
F-M std. error	(0.022)	(0.814)	(0.153)
# With predicted sign	42/48	48/48	45/48
Pooled regression	-0.150*** (0.037)		
Collapsed regression	-0.058** (0.022)		

2006 - 2017	RATIO	R&D to sales	ROE
Avg. coefficient	-0.037	0.001	0.063***
F-M std. error	(0.019)	(0.001)	(0.020)
# With predicted sign	9/12	12/12	10/12
Pooled regression	-0.037 (0.046)		
Collapsed regression	0.001 (0.034)		

4.6. Explaining growth controls

In the last part of my tests, I try to explain the adverse effect of future income growth control on the coefficient on RATIO. Hong, Kubik and Stein (2008) find that this variable has no effect on RATIO and is not statistically significant, whereas my results suggest exactly the opposite. In fact, my results suggest that the effect of future income growth is the largest of all controls. Remember, that the aggregate income in a region is the denominator of RATIO, so next I will also look at the numerator, aggregate book value in a region.

I run the baseline regression again, now adding the growth rate of aggregate book value in a region. I compute this variable in a similar way as the earlier growth controls, i.e. for years $t + 1$ to $t + 3$. Table 8 displays the results of this regression. Rows 1 to 3 three are the same as in Table 5 and rows 3 and 4 add the effect of region book value growth and the effect of both growth controls respectively. The effect of region book value growth on RATIO is smaller than the effect of region income growth, while still affecting the coefficient on RATIO more than other previous controls. Book value growth also has a positive and statistically significant effect on stock pricing.

As expected, adding both controls decrease the coefficient on RATIO even more and the individual effect of both controls is also decreased. Book value growth is still not able to take out the effect of income growth, so these two variables must proxy for some phenomenon that explains stock pricing. Considering the computation of RATIO, an increase in income should translate to increased risk tolerance of investors in a region, according to the CARA utility function. In this simple model, increased risk tolerance would in turn translate to increased investments in stocks and thus increasing the stock prices. While book value growth explains some variation in the coefficient on RATIO, the change in book value is probably more of a consequence than a cause for changes in stock prices.

The root of the difference in the results of income growth remain inconclusive. Having replicated RATIO and the controls as closely as possible, there are two possible explanations for why the results differ. First, Hong, Kubik and Stein (2008) might have done additional controlling on the data, that is not explicitly explained in their research or the variables are computed in a more complex way. The second, and more worrying, explanation is that the raw income data differs, resulting in a difference in the regression. As the raw values of RATIO are a close match, however, this seems unlikely.

Regardless of the difference in results, region income growth still doesn't wipe out the basic finding – the negative effect of RATIO on stock pricing. The negative sign still present and the coefficient staying statistically significant, a measurement error in one control doesn't affect the whole outcome of the tests.

Table 8 - Summary of RATIO constituent controls

	RATIO	Future region income growth	Future region bookvalue growth
1. Baseline specification	-0.183*** (0.041)		
2. Add region income growth	-0.123** (0.042)	2.47*** (0.213)	
3. Add region book value growth	-0.154*** (0.041)		0.801*** (0.064)
4. Add both future controls	-0.117** (0.042)	1.90*** (0.227)	0.569*** (0.068)

5. Conclusions

The findings of this thesis directly support the findings of Hong, Kubik and Stein (2008), from the replication of raw data to the replication of results. Theory suggests that local market conditions should affect stock pricing and indeed, this is confirmed in the tests. Ultimately, the differences between regions boils down to local supply and demand of stocks, which translates to differences in stock prices due to different equilibriums. The results are also robust, which is confirmed by three different regression approaches and additional controls. While some of the controls display a larger effect than expected, the basic findings still stay the same.

The results also confirm, that local bias treats large, well-known firms and smaller, local firms differently. Independent of the local market conditions, larger firms' stock prices tend to differ by less than 3% on average, between regions with highest and lowest values of RATIO. For smaller firms, the price difference could vary between 9% and 12%, depending on the proxy for visibility. This result is intuitive – a small firm, with geographically focused operations is more vulnerable to local market conditions and trading on such firms contains more information asymmetries, compared to large firms followed by many analysts.

Additionally, my tests with new data confirm, that the phenomenon persists to date and is economically significant. This finding is analogous to the findings of Levy and Levy (2014), who concluded at the time, that home bias had stayed pretty much unchanged for the last 15 years. Interestingly, the results suggest that the effect of local bias is in small decline, with the price discount implied by RATIO dropping by 1 percentage point. Even with these results, two questions remain – is the change in the pricing implications of RATIO permanent and if so, does the decline of the effect continue or does it converge to a certain level.

As a final note, it is good to understand what the results imply and what they don't. As Hong, Kubik and Stein (2008) point out, the model treats firms changing region with ceteris paribus assumption. This assumption is not exactly realistic, since the model treats firm ROE as a fixed variable and assumes it is independent from region change. Even more, this also implies that there is little benefit to be achieved from an arbitrage point of view, since the changes in returns are likely to be small relative to the risks and transaction costs implied by a region change (Hong, Kubik, & Stein, 2008).

References

Baltzer, M., Stolper, O., & Walter, A. (2013). Is local bias a cross-border phenomenon? Evidence from individual investors' international asset allocation. *Journal of Banking & Finance 37*, 2823-2835.

Coval, J. D., & Moskowitz, T. J. (1999). Home Bias at Home: Local Equity Preference in Domestic Portfolios. *The Journal of Finance 54*, 2045-2073.

Coval, J. D., & Moskowitz, T. J. (2001). The Geography of Investment: Informed Trading. *Journal of Political Economy 109*, 811-841.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy 81*, 607-636.

French, K. R., & Poterba, J. M. (1991). Investor Diversification and International Equity Markets. *Working paper no. 3609, National Bureau of Economic Research*.

Grinblatt, M., & Keloharju, M. (2001). How Distance, Language, and Culture Influence Stockholdings and Trades. *Journal of Finance 56*, 1053-1073.

Hietala, P. T. (1989). Asset Pricing in Partially Segmented Markets: Evidence from the Finnish Market. *Journal of Finance 44*, 697-718.

Hong, H., Kubik, J. D., & Stein, J. C. (2008). The only game in town: Stock-price consequences of local bias. *Journal of Financial Economics 90*, 20-37.

Huang, Y., Qiu, H., & Wu, Z. (2016). Local bias in investor attention: Evidence from China's Internet stock message boards. *Journal of Empirical Finance 38*, 338-354.

Ivković, Z., & Weisbenner, S. (2005). Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments. *Journal of Finance 60*, 267-306.

Jorion, P., & Schwartz, E. (1986). Integration vs. Segmentation in the Canadian Stock Market. *Journal of Finance 41*, 603-614.

Kumar, A., & Sulaeaman, J. (2015). Home away from Home: Geography of Information and Local Investors. *The Review of Financial Studies 28*, 2009-2049.

Levy, H., & Levy, M. (2014). The home bias is here to stay. *Journal of Banking & Finance 47*, 29-40.

Lindblom, T., Mavruk, T., & Sjögren, S. (2018). East or west, home is best: The birthplace bias of individual investors. *Journal of Banking & Finance 92*, 323-339.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703-708.

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies 22*, 435-480.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica 32*, 122-136.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory 13*, 341-360.

Seasholes, M. S., & Zhu, N. (2010). Individual Investors and Local Bias. *The Journal of Finance 65*, 1987-2010.

Tesar, L. L., & Werner, I. M. (1995). Home bias and high turnover. *Journal of International Money and Finance 14*, 467-492.